

Optimasi Decision Tree Dengan Particle Swarm Optimization Untuk Klasifikasi Prediksi Gender Pengguna E-Commerce

Oky Irnawati¹, Nia Nuraeni²

¹ Teknik Elektro, Universitas Bina Sarana Informatika

² Informatika, Universitas Nusa Mandiri

¹oky.okt@bsi.ac.id

²nia.nne@nusamandiri.ac.id

Abstrak— Belanja online telah menjadi gaya hidup saat ini. Hal ini berdampak pada banyaknya perusahaan aplikasi belanja online yang saling bersaing untuk menawarkan kemudahan dan kenyamanan pada pengguna aplikasi belanja online. Informasi mengenai pengguna akan sangat membantu pengusaha dibidang aplikasi belanja online sebagai pengetahuan untuk menerapkan strategi marketing guna terus mempertahankan pengguna aplikasi maupun merekrut pengguna aplikasi baru. Informasi tersebut termasuk mengenai gender pengguna yang dapat dimanfaatkan untuk teknik promosi, pemilihan brand ambassador, upgrade tampilan aplikasi dan lain sebagainya. Informasi mengenai gender pengguna aplikasi belanja online dapat di olah dengan menerapkan data mining menggunakan model klasifikasi dengan algoritma Decision Tree yang kemudian di optimasi dengan Particle Swarm Optimization (PSO). Hasil penelitian pada dataset e-commerce gender prediction menunjukkan adanya peningkatan akurasi Decision Tree yang dioptimasi Particle Swarm Optimization (PSO) sebesar 0,15%, tingkat akurasi yang sebelumnya sebesar 85,22% dengan angka kurva ROC 0.759 menjadi 85,37% dan kurva ROC 0.764 termasuk dalam fair classification.

Kata Kunci— Data Mining, Decision Tree, Particle Swarm Optimization (PSO), E-commerce Gender Prediction.

I. PENDAHULUAN

Penggunaan teknologi internet yang semakin meningkat telah merubah gaya berbelanja saat ini menjadi online. Trend berbelanja online melalui aplikasi e-commerce semakin hari semakin meningkat sehingga bermunculan aplikasi e-commerce yang saling bersaing menawarkan kemudahan bertransaksi, kenyamanan, dan keamanan dalam berbelanja secara online untuk mendapatkan kepercayaan konsumen. Memiliki pengetahuan mengenai pelanggan melalui aktivitas yang tercatat dalam aplikasi e-commerce dapat menjadi informasi yang berharga dalam memenangkan persaingan antar aplikasi e-commerce.

Menurut penelitian Oh, dkk berjudul “Age- and Gender-Based Market Segmentation”, usia, jenis kelamin, lokasi, atau pendapatan menjadi segmentasi pelanggan aplikasi e-commerce [1][2]. Banyak studi penelitian pemasaran mengenai perbedaan faktor pengambilan keputusan, frekuensi dan waktu tinggal berbelanja online yang dipengaruhi oleh perbedaan gender [3]. Segmentasi pelanggan ini memiliki peranan penting untuk mengidentifikasi kebutuhan pelanggan

yang nantinya dijadikan rujukan untuk membuat keputusan strategi bisnis [4].

Penerapan data mining berdasarkan segmentasi pelanggan e-commerce dapat digunakan untuk menjalankan rencana strategis aplikasi e-commerce dalam meningkatkan pelanggan selaku pemakai aplikasi [5][6]. Informasi mengenai segmentasi dan kebiasaan pelanggan e-commerce dapat menghasilkan rekomendasi secara personal, voucher diskon yang tepat, maupun pemilihan brand ambassador sebagai media promosi dan marketing untuk mempertahankan pelanggan atau dapat juga menarik pelanggan aplikasi e-commerce pesaing [7]. Segmentasi pelanggan berdasarkan gender dapat diprediksi menggunakan data mining klasifikasi dengan target variabel kategori sebagai class berdasarkan data training [8].

Decision Tree adalah salah satu algoritma dalam metode klasifikasi dengan kemudahan interpretasi hasil yang dimilikinya, memiliki alur yang dapat ditelusuri dari simpul ke akar simpul untuk mengetahui prediksi kelas [9]. Algoritma Decision Tree bergantung pada pembuatan desain pohon, jika ada atribut yang tidak sesuai maka desain pohon menjadi lemah [10]. Hal ini yang mendasari perlunya optimasi dengan Particle Swarm Optimization (PSO) untuk peningkatan akurasi. Terinspirasi dari perilaku hewan dalam mencari mangsa, PSO terdiri dari sekumpulan partikel dengan posisi terbaik optimasi dalam ruang fitur sehingga PSO memiliki kelebihan menjadi lebih efisien [11].

II. KAJIAN PUSTAKA

Penelitian mengenai pengaruh gender terhadap perilaku berbelanja online telah dibahas pada penelitian sebelumnya dan menghasilkan kesimpulan gender memiliki pengaruh namun tidak signifikan, implementasi machine learning dengan optimasi support vector machines (SVMs) meningkatkan performa sistem dalam memprediksi gender pengguna e-commerce [12][13]. Beberapa penelitian terdahulu mengenai metode klasifikasi diantaranya penelitian yang dilakukan oleh Pratiwi dan Arifin berjudul Optimasi Decision Tree Menggunakan Particle Swarm Optimization untuk Klasifikasi Kesuburan pada Pria menghasilkan kesimpulan nilai akurasi pada fertility dataset mengalami peningkatan dengan algoritma Decision Tree yang dikombinasikan dengan Particle Swarm Optimization (PSO) sebesar 0,33%. Nilai akurasi 90,00% dan AUC 0,769 dengan

Penelitian-penelitian tersebut menjadi dasar pada penelitian ini untuk menerapkan metode Decision tree yang dioptimasi dengan Particle Swarm Optimization (PSO) untuk dataset e-commerce gender prediction.

Tujuan penelitian adalah untuk menguji algoritma decision tree yang dioptimasi dengan Particle Swarm Optimization (PSO) dalam menghasilkan pola klasifikasi pada data dari repository keagle untuk menentukan gender pengguna e-commerce berdasarkan waktu mulai penggunaan aplikasi, waktu selesai penggunaan aplikasi serta produk yang dilihat. Informasi mengenai gender tersebut dapat digunakan untuk menentukan strategi marketing seperti penentuan brand ambassador, promosi melalui diskon dan voucher serta rekomendasi produk untuk meningkatkan pengguna aplikasi e-commerce.

Modeling berarti menggunakan algoritma untuk membuat pola. Pola dalam penelitian ini adalah klasifikasi dengan algoritma klasifikasi yang akan diuji yaitu Decision Tree yang

kemudian dioptimasi dengan Particle Swarm Optimization (PSO). Menggunakan tools rapidminer untuk mendapatkan pola algoritma Decision Tree dan nilai akurasi hanya dengan Decision Tree dan nilai akurasi setelah dioptimasi.

5. Evaluation

Evaluation untuk menentukan akurasi dan efektivitas dari modelling. Pada penelitian model klasifikasi nilai akurasi diambil dari data testing, performance atau efektivitas disajikan dengan dalam bentuk Area Under Curve (AUC) dengan klasifikasi range 0.90-1.00 untuk excellent classification, 0.80-0.90 untuk good classification, 0.70-0.80 untuk fair classification, 0.60-0.70 untuk poor classification dan 0.50-0.60 untuk failure.

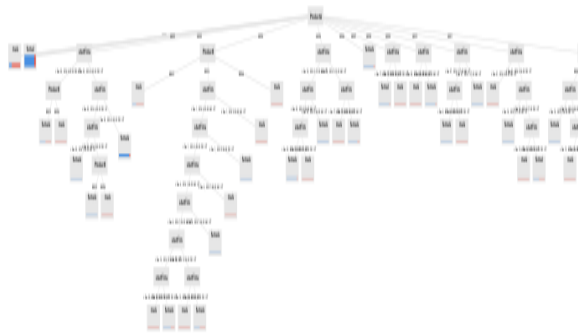
6. Deployment

Hasil akurasi terbaik yang didapatkan dari tahap evaluasi akan menentukan algoritma terbaik yang dapat diterapkan pada dataset e-commerce gender prediction.

IV. HASIL DAN PEMBAHASAN

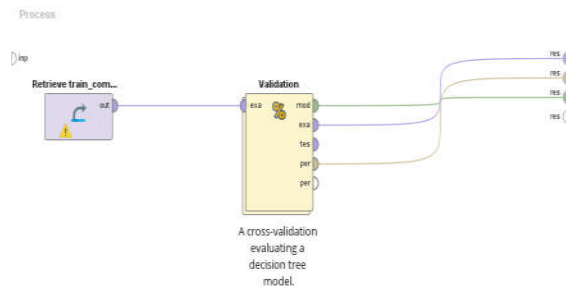
A. Decision Tree

Menggunakan tools rapid miner untuk menemukan pola dengan algoritma klasifikasi decision tree didapatkan pola pada gambar 2.



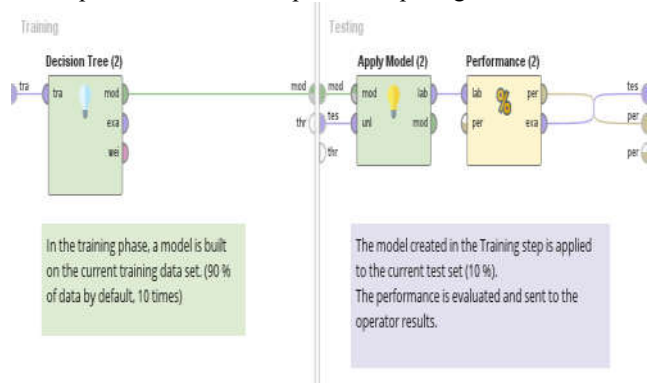
Gbr 2. Pola Decision Tree

Kemudian dilakukan cross validation dengan menggunakan ten cross validation yaitu melakukan percobaan pada partisi data testing dan data training sebanyak 10 kali untuk mendapatkan data testing yang lebih stabil.



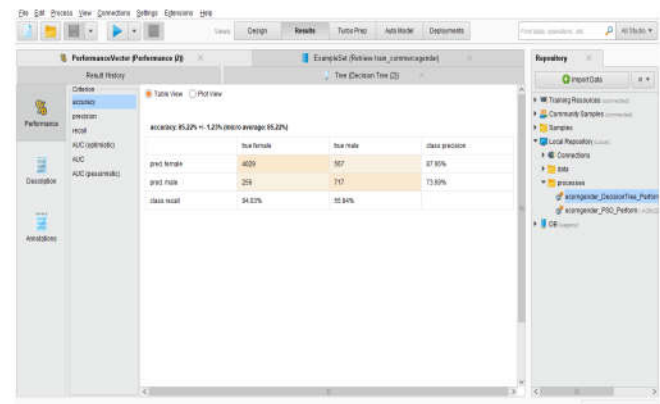
Gbr 3. Implementasi Cross-Validation Decision Tree

Data yang telah divalidasi diterapkan dengan algoritma Decision Tree untuk melihat performa dari data testing. Penerapan Decision Tree dapat dilihat pada gambar 4.



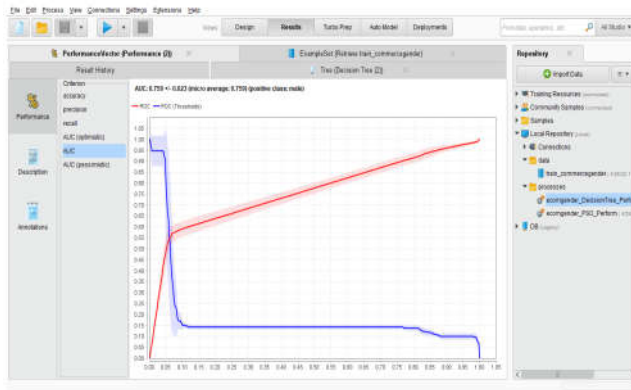
Gbr 4. Implementasi Performance Decision Tree

Hasil akurasi algoritma Decision Tree pada data testing e-commerce gender sebesar 85.22% dapat dilihat pada gambar 5.



Gbr 5. Hasil Akurasi Decision Tree

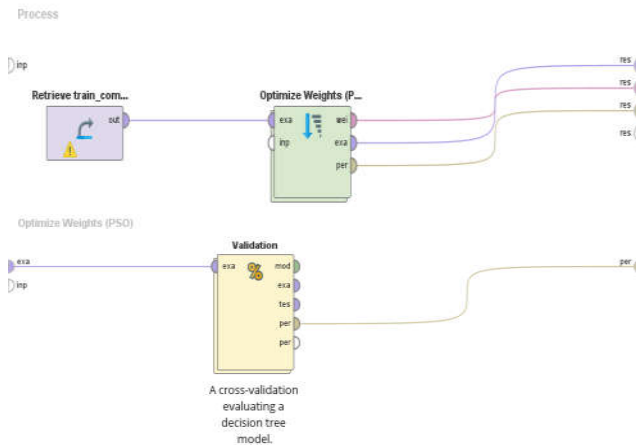
Kurva ROC menunjukkan keakuratan algoritma Decision Tree pada data testing e-commerce gender. Hasil pada gambar 6 menunjukkan hasil grafik AUC sebesar 0.759.



Gbr 6. Grafik ROC Decision Tree

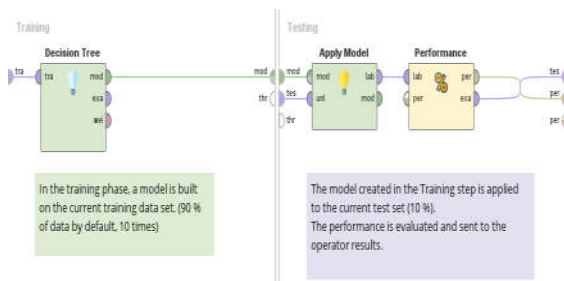
B. Particle Swarm Optimization (PSO)

Berdasarkan pengujian sebelumnya dilakukan optimasi dengan metode Particle Swarm Optimization (PSO) seperti pada gambar 7.



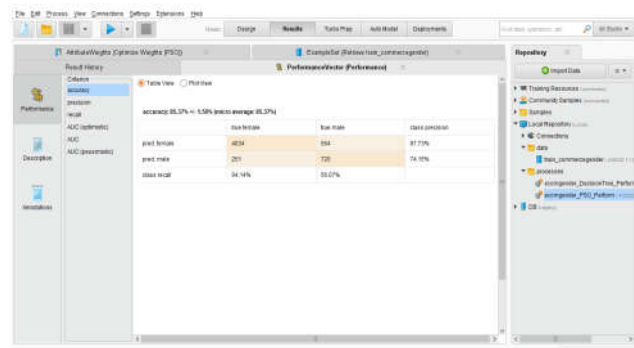
Gambar 7. Implementasi Cross-Validation PSO

Kemudian dilakukan pengujian performa pada data testing dengan algoritma Decision Tree yang dioptimasi dengan metode Particle Swarm Optimization (PSO) tersebut.



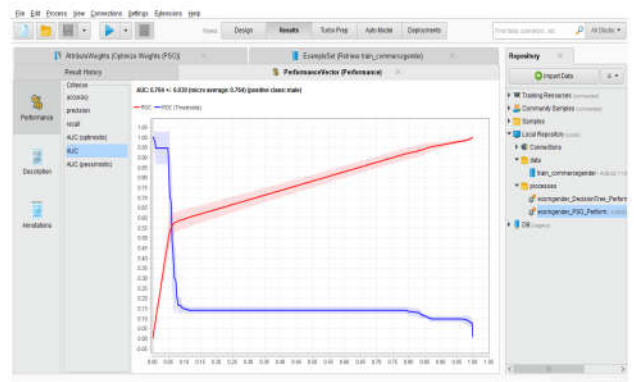
Gbr 8. Implementasi Performance PSO

Akurasi dengan menerapkan optimasi mengalami peningkatan dengan hasil yang didapatkan sebesar 85.37% seperti yang terlihat pada gambar 9.



Gbr 9. Hasil Akurasi PSO

Kurva ROC setelah algoritma Decision Tree dioptimasi menunjukkan peningkatan keakuratan dengan hasil 0.764 seperti terlihat pada gambar 10.



Gbr 10. Grafik ROC PSO

V. KESIMPULAN

Kesimpulan dari hasil penelitian yang dilakukan pada dataset e-commerce gender prediction dengan metode klasifikasi menggunakan algoritma Decision Tree dihasilkan tingkat akurasi sebesar 85.22% dengan angka kurva ROC 0.759. Peningkatan akurasi menggunakan metode Particle Swarm Optimization (PSO) sebesar 0.15%. Tingkat akurasi dengan Particle Swarm Optimization (PSO) menjadi 85,37% dan kurva ROC 0.764 termasuk dalam fair classification. Prediksi gender dari pengguna aplikasi e-commerce dibutuhkan untuk strategi marketing sehingga pada penelitian selanjutnya dapat menggunakan kumpulan data e-commerce lain dengan metode yang ditingkatkan.

REFERENSI

- [1] H. Oh, S. C. Parks, and F. J. DeMicco, "Age- and Gender-Based Market Segmentation," *Int. J. Hosp. Tour. Adm.*, vol. 3, no. 1, pp. 1–20, 2002.
- [2] V. Urbancokova, M. Kompan, Z. Trebulova, and M. Bielikova, "Behavior-based customer demography prediction in e-commerce," *J. Electron. Commer. Res.*, vol. 21, no. 2, pp. 96–112, 2020.
- [3] Z. Zhang, H. Li, F. Meng, and S. Qiao, "Gender Difference In Restaurant Online Booking Timing And The Moderating Effects Of Sell-Out Risk And Information Type," *J. Electron. Commer. Res.*, vol. 19, no. 3, pp. 266–279, 2018.
- [4] S. Koul and T. M. Philip, "Customer Segmentation Techniques on E-Commerce," in *2021 International Conference on Advance Computing and Innovative Technologies in Engineering, ICACITE 2021*, 2021, vol. 7.
- [5] O. Nurdianawan and R. Suryatana, "Implementasi Strategis Penjualan Fashion Melalui Penerapan Data Mining Untuk Pengelompokan Market Share Fashion," *J. Tek. Komput. AMIK BSI*, vol. V, no. 1, pp. 97–104, 2019.
- [6] R. Punhani, V. P. S. Arora, A. S. Sabitha, and V. K. Shukla, "Segmenting E-Commerce Customer through Data Mining Techniques," in *Journal of Physics: Conference Series CONSILO 2020*, 2021, vol. 1714, no. 1.
- [7] S. Prasetyo, Sandy, and A. R. Hakim, "Penerapan Fuzzy C-Means Kluster Untuk Segmentasi Pelanggan E-Commerce Dengan Metode Recency Frequency Monetary (RFM)," *J. GAUSSIAN*, vol. 9, no. 4, pp. 421–433, 2020.
- [8] Y. Mardi, "Data Mining : Klasifikasi Menggunakan Algoritma C4.5," *Edik Inform.*, vol. 2, no. 2, pp. 213–219, 2017.
- [9] A. Amrulloh and A. P. Wibowo, "Implementasi Algoritma Decision Tree Untuk Mengklasifikasi Kondisi Kesuburan Pria," *JASIEK (Jurnal Apl. Sains, Informasi, Elektron. dan Komputer)*, vol. 1, no. 1, pp. 7–11, 2019.
- [10] T. W. Pratiwi and T. Arifin, "Optimasi Decision Tree Menggunakan Particle Swarm Optimization untuk Klasifikasi Kesuburan pada Pria," *Sist. (Jurnal Sist. Informasi)*, vol. 10, no. 1, pp. 1–12, 2021.
- [11] T. Arifin, "Implementasi Algoritma PSO Dan Teknik Bagging Untuk Klasifikasi Sel Pap Smear," *J. Inform.*, vol. 4, no. 2, pp. 155–162, 2017.
- [12] A. G. Purwaningsih and Nurhadi, "Pengaruh Promosi Penjualan Dan Gender Terhadap Perilaku Impulse Buying Pada E-Commerce Shopee," *J. Ilm. STIE MDP*, vol. 10, no. 2, pp. 159–167, 2021.
- [13] M. M. Khan, M. G. Sohrab, and M. A. Yousuf, "Customer gender prediction system on hierarchical E-commerce data," *Beni-Suef Univ. J. Basic Appl. Sci.*, vol. 9, no. 1, 2020.
- [14] B. Altunan, E. D. Arslan, M. Seyis, and M. Birer, "A Data Mining Approach to Predict E-Commerce Customer Behaviour," in *Proceedings of the International Symposium for Production Research 2018*, 2019, pp. 29–43.
- [15] C. Wang and Y. Ji, "Gender Prediction of Consumers Using Offline Purchase Data," vol. 150, no. Macmc 2017, pp. 281–290, 2018.
- [16] Z. Ge, Z. Song, S. X. Ding, and B. Huang, "Data Mining and Analytics in the Process Industry: The Role of Machine Learning," *IEEE Access*, vol. 5, pp. 20590–20616, 2017.